Relationship between global suicide rates and gross domestic product (GDP)

Name : Yi Xu

Student ID : 28516222

Tutor : Jeffery Liu

Table of Contents

1. Introduction ... 3
  1.1 Background ... 3
  1.2 Research question ... 3
  1.2.1 Question list ... 3
  1.2.2 Question description and motivation ... 3
2.Data wrangling ... 3
  2.1 Description of data sources ... 3
  2.2 Data wrangling ... 4
3. Data checking ... 6
4.Data Exploration ... 8
  4.1 Global analysis ... 8
  4.2 Relationship between economic factors and suicide rate ... 9
  4.3 Proportion of male and female suicide rates ... 11
5.Conclusion ... 12
6.Reflection ... 12
7.Bibliography ... 13

1. Introduction

1.1 Background

Suicide is a major global health organization problem. According to statistics from the World Health Organization, about 1 million people committed suicide each year, and 10-20 million people with attempted suicide (Fleischmann and Leo, 2014). The National Institute of Mental Health (NIMH) estimates that 6.7 percent of the American adult had an anxiety disorders(Healthline, n.d., para. 3). In recent years, due to economic and mental pressure, the suicide rate has become the focus in nowadays. This report will explore whether economic changes have an impact on suicide rates.

1.2 Research question

1.2.1 Question list

1.Which country has the highest suicide rate?

2. How changes in the global economy will affect suicide rates ?

3. The impact of economic changes on the suicide rate of different genders

1.2.2 Question description and motivation

1. The first question can analyze which country has the highest suicide rate in the world. After finding the country, it can be provided to scholars who are interested in high suicide rates. This can help scholars find out the causes and solutions of high suicide rates and provide some reference value for other countries.

2. There are many factors for suicide. This report will select the Gross Domestic Product (GDP) to explore the correlation between economic changes and the global suicide rate in the four years(2000, 2010, 2015, 2016). This question can help us understand whether the improvement of economic level can improve people's happiness and reduce suicide rate.

3. This question explores the influence of gender and economic factors on suicide rate and finds the relationship between gender and suicide rate at the same economic level. This finding can help us to observe whether male and female will cause different suicide rates in the context of economic changes.

2.Data wrangling

2.1 Description of data sources

a. Global suicide rate estimates which is contains crude estimates rate, crude suicide rates per gender and country from 2000 to 2016.

b. World development indicators: GDP (current US$) by country from 1960 -2018

The first data source can help me answer to question 1. Combining data resources, a and b can help me answer to question 2 and 3.

2.2 Data wrangling

I use python language for data wrangling. The following are the data wrangling steps:

(1) suicide rate data wrangling

1.Read the suicide rate data in python and show the data head

2. Change the column name(Figure 1)

3. Convert columns(2016,2015,2010,2000) into rows

4. Order the data frame by column(Country and Year)

5. Change the order of column

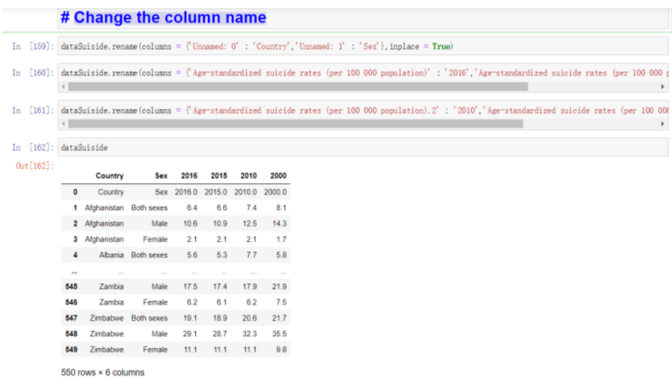


Figure 1 Change the column name

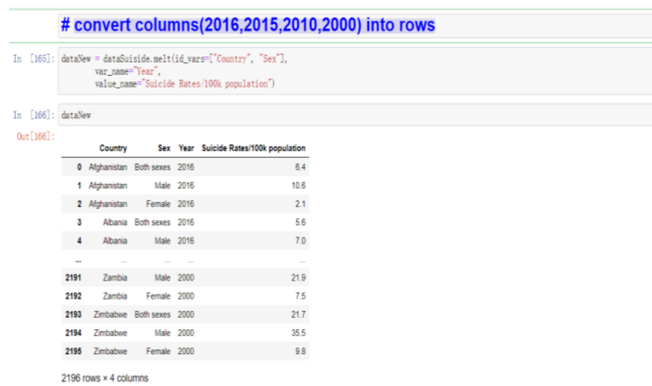


Figure 2 Convert columns into rows

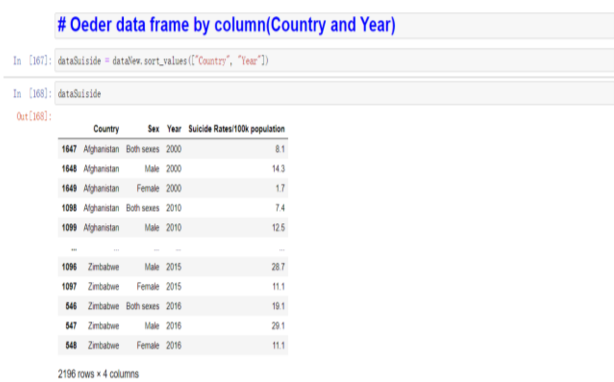


Figure 3 Order the data frame by column

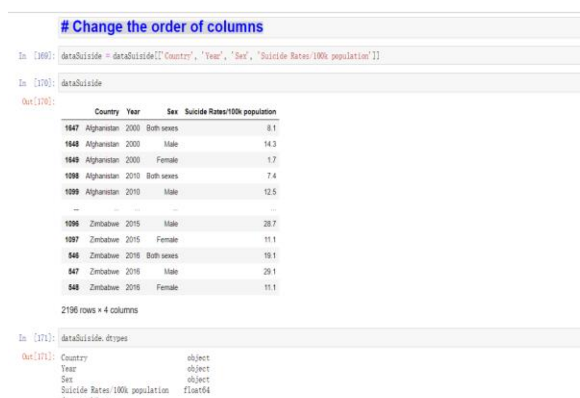
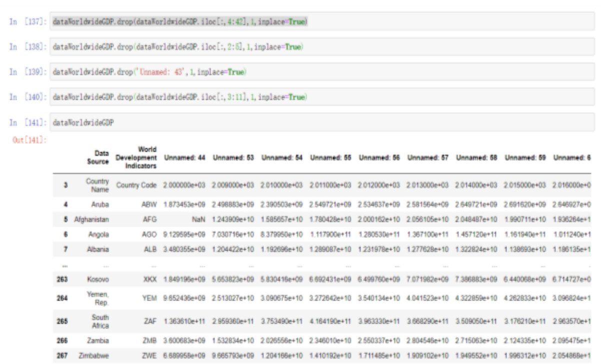


Figure 4 Change the order of column

(2) GDP data wrangling

1. Read the GDP data in python and show the data head
2. Delete unwanted rows(Figure 5)
3. Change the column name(Figure 6)
4. Convert columns(2010-2018) into rows(Figure 7)
5. Order the data frame by column(Country and Year) (Figure 8)



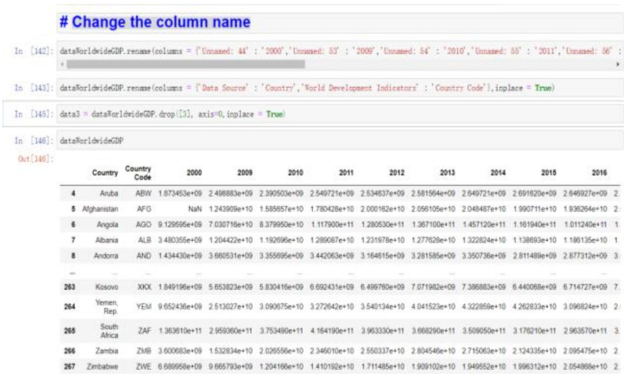
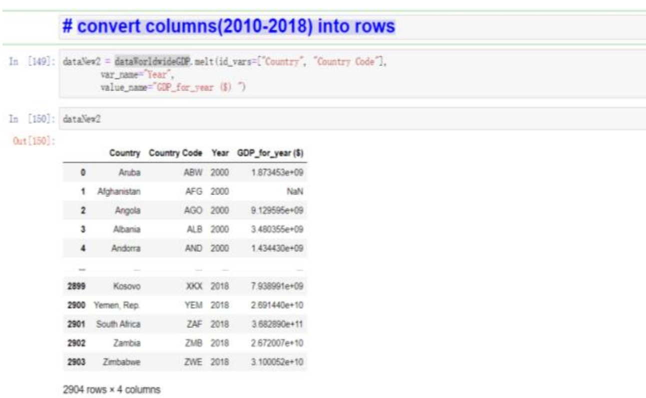


Figure 5 Delete unwanted rows

Figure 6 Change the column name



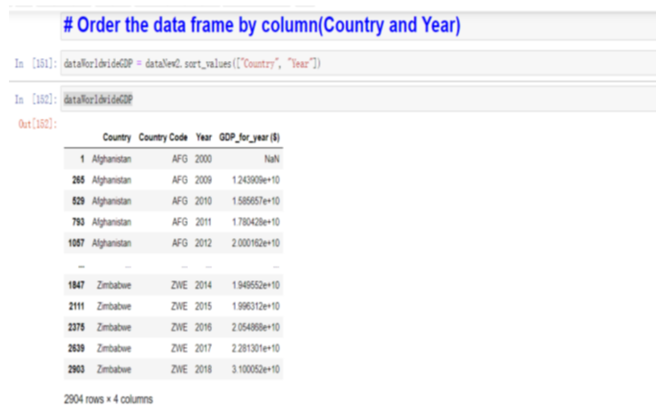


Figure 7 Convert columns into rows

Figure 8 Order the data frame by column

(3) Combine two data frames using “Country” and “Year”

Figure 9 Combine two data frames

3. Data checking

I use python language for data checking. The following are the data checking steps:

1. Check the null value (Figure 10)

2. Addressing null value(Figure 11,12)

Since some countries do not have GDP values, I choose to check the values and insert them in the data frame

3. Check null value again(Figure 13)

Due to the garbled name of this country ("C ?? te d'Ivoire"), I chose to delete this country's data.

4. Find duplicate records(Figure 14)

Since there are no duplicate rows in my data, the data does not need to be processed.

5.Data logic check(Figure 15)

Check whether the values of suicide rate and GDP are less than 0.

5. Check data outliers(Figure 16,17,18)

I used Scatter and Boxplot graph to explore abnormal data(Figure 16&17). The data shows that some countries have abnormally high suicide rates, so I chose to select countries with suicide rates greater than 60 (Figure 18). Although there are some ways to deal with these outliers, such as removing or averaging these data, but in order to show the real data, the outliers will be intentionally saved in the data source.

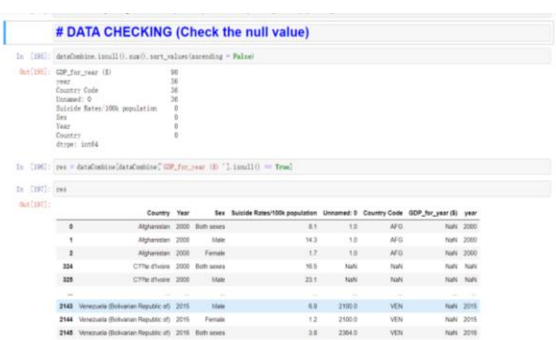


Figure 10 Check the null value

Figure 11 Addressing null value

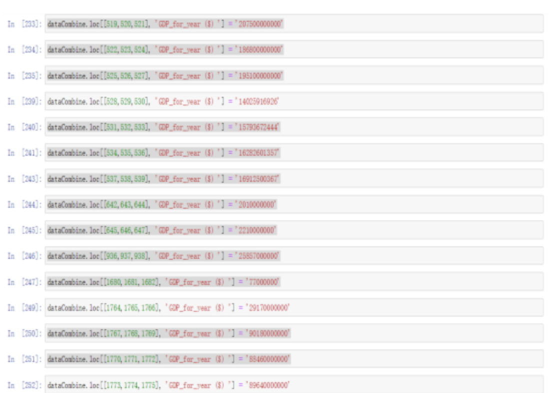


Figure 12 Addressing null value

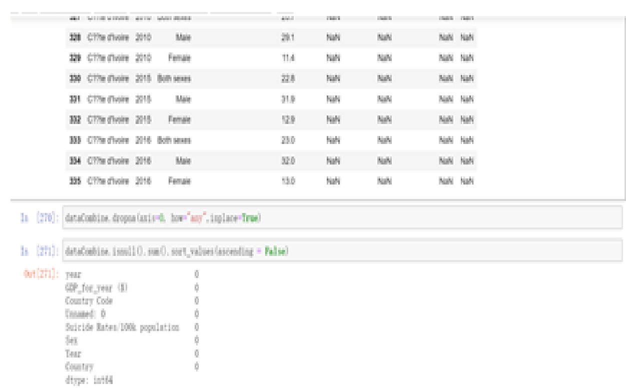


Figure 13 Check null value again

DATACHECKING(Check for duplicates)

```
In [17]: dup = dataCombine[dataCombine.duplicated()]

In [18]: dup
Out[18]:
```

Unnamed: 0	Country	Year	Sex	Suicide Rates/100k population	Unnamed: 0.1	Country Code	GDP_for_year ($)	year

Figure 14 Find duplicate records

DATACHECKING(Check if the data is logical)

```
In [28]: SuicideLogical = dataCombine[dataCombine.loc[:, 'Suicide Rates/100k population'] < 0]

In [29]: SuicideLogical
Out[29]:
```

Country	Country Code	Year	Sex	Suicide Rates/100k population	GDP_for_year ($)

```
In [30]: GDPLogical = dataCombine[dataCombine.loc[:, 'GDP_for_year ($) '] < 0]

In [31]: GDPLogical
Out[31]:
```

Country	Country Code	Year	Sex	Suicide Rates/100k population	GDP_for_year ($)

Figure 15 Data logic check

DATACHECKING(Check data outliers)

```
In [287]: import matplotlib.pyplot as plt

In [288]: plt.scatter(x=dataCombine['Year'],y=dataCombine['Suicide Rates/100k population'])
Out[288]: <matplotlib.collections.PathCollection at 0x26a4d970588>
```

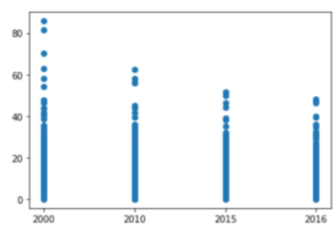


Figure 16 Check data outliers

```
suicideAndGDP %>% boxplot(Suicide.Rates.100k.population ~ Year,
                          data = ., main="Boxplot of Suicide Rate",
                          ylab = "Suicide Rates per 100k populaction",
                          xlab = "Year", col="Yellow")
grid()
```

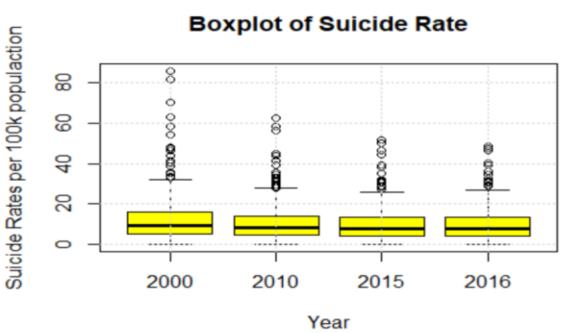


Figure 17 Check data outliers

```
In [289]: dataCombine2 = dataCombine[dataCombine.loc[:, 'Suicide Rates/100k population'] > 60]

In [290]: dataCombine2
Out[290]:
```

	Country	Country Code	Year	Sex	Suicide Rates/100k population	GDP_for_year ($)
169	Belarus	BLR	2000	Male	70.4	1.27369e+10
1021	Kazakhstan	KAZ	2000	Male	63.1	1.8292e+10
1153	Lithuania	LTU	2000	Male	81.7	1.15392e+10
1621	Russian Federation	RUS	2000	Male	85.8	2.5971e+11
1624	Russian Federation	RUS	2010	Male	62.3	1.52492e+12

Figure 18 suicide rate greater than 60

4.Data Exploration

4.1 Global analysis

Using bar charts in R to analyze the values of suicide rates in all countries in 2000, 2010, 2015, and 2016 (Figure 19&20). Due to the large number of countries, I arranged the values of different suicide rates in descending order. Four colors (red, green, blue, purple) are used in the chart to represent Years(2000, 2010, 2015, 2016). It can be found that the top three countries with high suicide rate are Russian Federation (RUS), Lithuania (LTU) and Guyana (GUY), but the country of Grenada (GRD), Barbados (BRB), Antigua and Barbuda (ATG) are with the low level of the suicide rate(Figure 21&22).

```r
data2 <- data %>% mutate(Country.Code = fct_reorder(Country.Code, desc(Suicide.Rates.100k.population)))

  ggplot(data2, aes(x = Country.Code, y =Suicide.Rates.100k.population, fill=factor(Year))) +
    geom_bar(position = "dodge", stat = "identity") +
    scale_size(range = c(.1, 10), name="Suicide.Rates.100k")+
  scale_fill_viridis(discrete=TRUE, guide=FALSE, option="A") +
    theme_ipsum() +
    theme(legend.position="bottom") +
    labs(title = "Global suicides per 100k, by Country",
         x = "Country code",
         y = "Suicides per 100k")+
    theme(legend.position = "none")
```

Figure 19 Global suicide rate/100k by country (code)

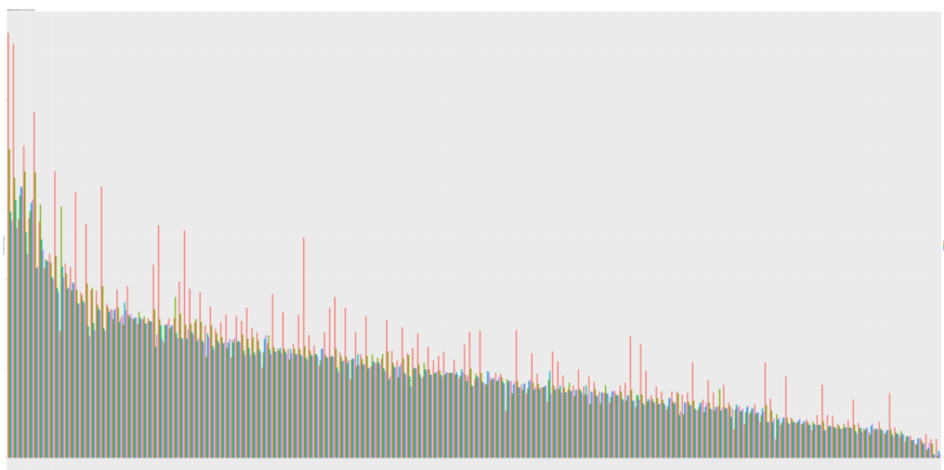
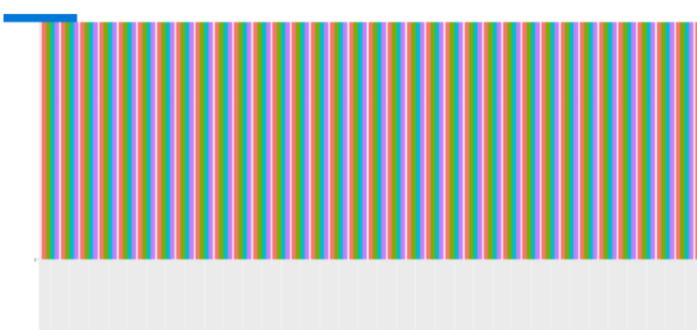


Figure 20 Global suicide rate/100k by country



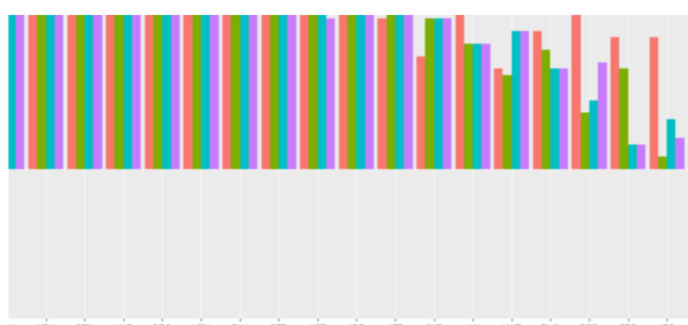


Figure 21 Countries with high suicide rate ***Figure 22 Countries with low suicide rate***

4.2 Relationship between economic factors and suicide rate

When studying this problem, the data of year, GDP and suicide rate were selected to measure the relationship between GDP and suicide rate. The line chart shows the change of suicide rate in four years under different economic conditions (Figure 23&24). In the line chart, the abscissa is the GDP value, and the ordinate is the suicide rate/100k, and different colors are used to indicate different years.

It can be found that when the GDP is at a low level, the data is too dense to observe a significant trend, but when the GDP reaches about 5.0e+12(e means "multiply by 10"), the suicide rate has dropped significantly. Interestingly, when GDP reached 1.0e+13, the suicide rate increased significantly.

```
ggplot(data1,
       aes(x = GDP,
           y = Suicide.Rates.100k.population, colour = factor(Year))) +
  geom_line(alpha=0.5) +
  ggtitle("Plot of Yearly Suicide Rate and GDP per Capital")+
  scale_size(range = c(.1, 10), name="Suicide.Rates.100k")+
  scale_fill_viridis(discrete=TRUE, guide=FALSE, option="A") +
  theme_ipsum() +
  scale_colour_wsj()
```

Figure 23 Plot of yearly suicide rate and GDP per capital(code)

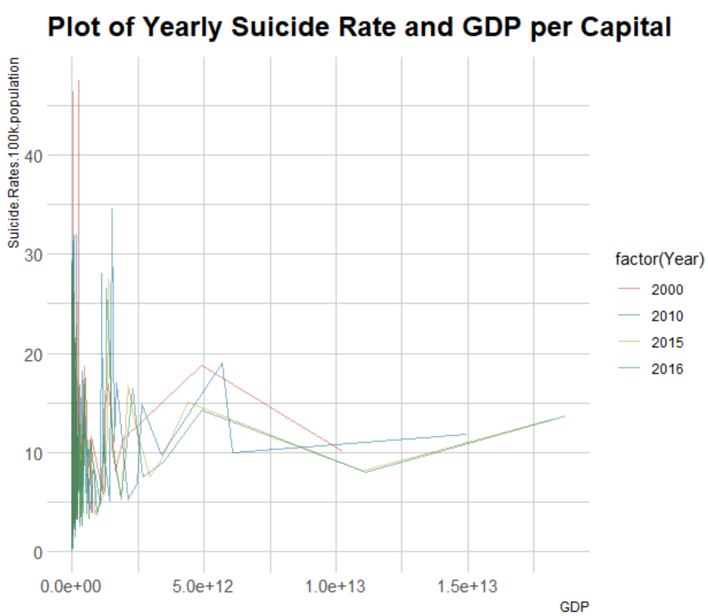


Figure 24 Plot of yearly suicide rate and GDP per capital

In order to tested the hypothesis that the suicide rate tends to increase when GDP is at a relatively high level, I created correlation to explore the correlation coefficient between GDP and suicide rate(Figure 25&26). It can be found that the correlation coefficient between GDP and the suicide rate is 0.03. In addition,the data trends of scatterplots also reflects this phenomenon(Figure 27&28).The horizontal and vertical coordinates in the plot represent GDP and suicide rate, and use different colors to correspond to different years. Each country represents a data point. The graph shows that there is a weak but positive linear relationship between GDP and suicide rate in different years. This means that the wealthier countries may have a higher proportion of suicide rates.

```
library(psych)
library(corrplot)
data <- read.csv("bothsexmanyyear.csv")
data<-data[,-1]
data<-data[,-1]
data<-data[,-1]
data<-data[,-1]
data<-data[,-1]
cormat<-corr.test(data)
cormat
corrplot(cormat$r)
```

Figure 25 correlation coefficient (code)

```
Call:corr.test(x = data)
Correlation matrix
                             Suicide.Rates.100k.population  GDP
Suicide.Rates.100k.population                         1.00 0.03
GDP                                                   0.03 1.00
Sample Size
[1] 720
Probability values (Entries above the diagonal are adjusted for multiple tests.)
                             Suicide.Rates.100k.population  GDP
Suicide.Rates.100k.population                         0.00 0.38
GDP                                                   0.38 0.00

 To see confidence intervals of the correlations, print with the short=FALSE option
```

Figure 26 correlation coefficient

```
ggplot(data2, aes(x = GDP, y = Suicide.Rates.100k.population,
                  colour  = as.factor(Year)))+
  geom_point() +
  scale_colour_brewer(palette = "Set2") +
  geom_smooth(method = "lm")+
  scale_x_continuous(labels=scales::dollar_format(prefix="$"))+
  labs(title = "Correlation between GDP (per capita) and Suicides per 100k",
     x = "GDP (per capita)",
     y = "Suicides per 100k") +
  theme(legend.position = "top")+
  theme_ipsum()+
  guides(colour=guide_legend(title="Year"))
```

Figure 27 Correlation between GDP and suicide rate(code)

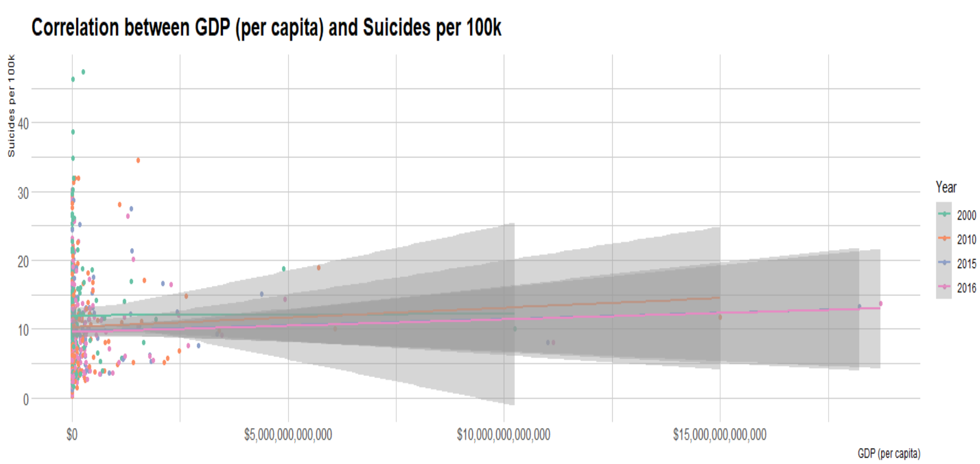


Figure 28 Correlation between GDP and suicide rate

4.3 Proportion of male and female suicide rates

The density map will be used to explore the ratio of male and female to suicide(Figure 29&30). The abscissa of the graph represents the proportion of suicidal groups per 100,000 people, and the ordinate represents the intensity of the suicide rate. The green part of the graph represents the trend of female suicide rate concentration, and the purple part represents males. By observing the graph, it can be seen that the female suicide rate peaks at around 3(per 100k population), about 0.16(density), and the male peaks around 13(per 100k population), which is about 0.045(density). This means that the suicide rate of female mostly keeps at about 3 / 100,000 ,but the suicide rate of male is about 13 / 100,000. The graph can indicate that the suicide rate of female is higher than the male.

```
ggplot(data=data1, aes(x=Suicide.Rates.100k.population, group=Sex, fill=Sex)) +
geom_density(position = "stack", alpha = 0.6) +
  scale_fill_viridis(discrete=TRUE) +
  scale_color_viridis(discrete=TRUE) +
  theme_bw() +
  scale_fill_brewer(palette="Accent")+
  ylab("Density") +
  xlab("Suicide Rates per 100k populaction")+
  ggtitle("Density plot of suicide rates by sex")+
guides(fill=guide_legend(title="Sex"))
```

Figure 29 Density plot of suicide rate by sex(code)

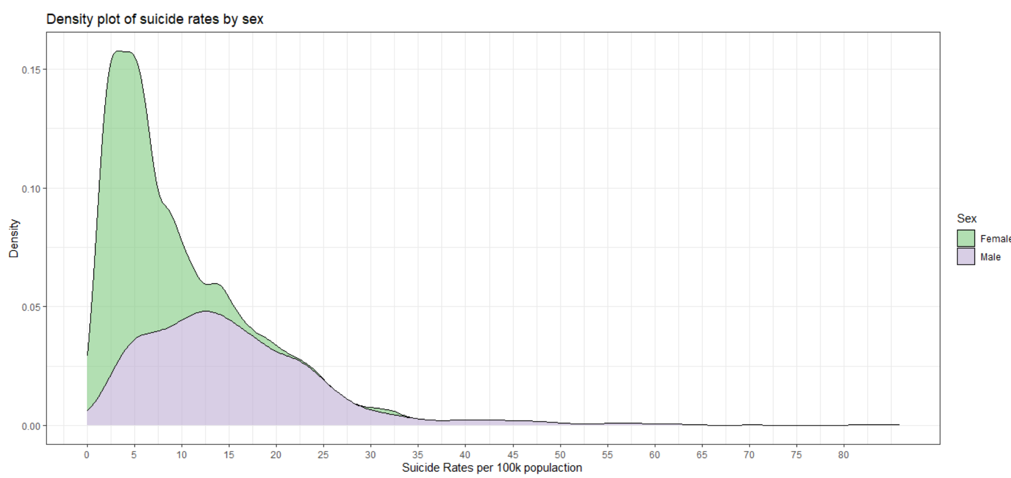


Figure 30 Density plot of suicide rate by sex

In order to find the relationship between the gender suicide rate by GDP and year, I created the scatterplot to analyse it(Figure 31&32). The abscissa represents GDP growth, and the ordinate represents different genders. The graph uses different colors to show different years, and circles with different radii to show the different level of suicide rate.

In this scatterplot, it can be observed that some countries with high level of GDP, the suicide rates of female has a slight decrease, but the suicide rate of male increased slightly. This may reflect that the wealthier countries the the suicide rate of male may increase, but the the suicide rate of female may decrease. In addition, It can be observed in the graph that no matter at what stage of GDP or in what year, the gray circle radius of male is always greater than that of female, which means that the the suicide rate of male is always significantly higher than that of female.

```r
mytheme <- theme_minimal()+
  theme(
    panel.grid.major.y=element_blank(),
    panel.grid.minor.y=element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1),plot.title=element_text(hjust =0.5),
    axis.line.y=element_line(linetype=1,color='grey'),axis.line.x=element_line(linetype=1,color='grey'),
    axis.ticks = element_line(linetype=2,color='grey'),panel.grid=element_line(linetype=2,color='grey'),
    legend.background = element_rect(fill="gray90", size=0,color='white'),legend.text=element_text(face="bold",size=8),
    legend.title=element_text(face="bold",size=8),axis.text=element_text(face="bold",size=8)
  )

 ggplot(data1, aes(x=GDP_for_year,
                  y=Sex,size = Suicide.Rates.100k.population,
                  color=factor(Year))) +
  geom_point(alpha=0.4,show.legend = TRUE)+
  labs(x='GDP',y='Sex(F/M)',title='Suicide rate of different gender by year and GDP')+
  scale_size(range = c(.1, 10), name="Suicide.Rates.100k")+
  scale_fill_viridis(discrete=TRUE, guide=FALSE, option="A") +
  theme_ipsum() +
  theme(legend.position="bottom") +
  theme(legend.position = "none")+
  mytheme+
   scale_x_continuous(breaks=seq(0, 2.0e+13, 0.2e+13))+
   scale_size_continuous(breaks=seq(0, 80, 10))+
   guides(color=guide_legend(title="Year"))+
   scale_size_continuous(range=c(2,30))
```

Figure 31 Suicide rate of different gender by year and GDP(code)

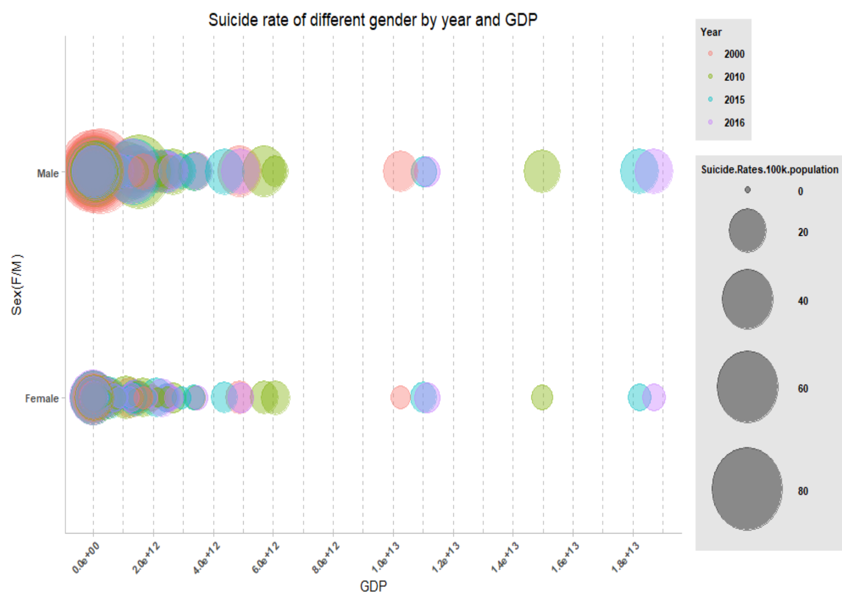


Figure 32 Suicide rate of different gender by year and GDP

5.Conclusion

This report explores the relationship between suicide rates, gender and economic factors in different countries around the world. It can be found that the country with the highest suicide rate in the world is the Russian Federation (RUS). In addition, economic changes have weakly positive linear relationship with the suicide rate. It means that countries with high level of GDP, the suicide rate is likely to be higher. Finally, the suicide rate of male is generally higher than female suicide rates. In the process of GDP growth, the suicide rate of female decreased, but males increased slightly.

6.Reflection

As for this project, I learned how to use python to process data, and using R language to draw different visualizations. In this project, I found that my data source should increase the latitude and longitude data of each country, so that I can use different type of maps to display the different level of the global suicide rate in the world.

Bibliography

Fleischmann, A., & De Leo, D. (2014). The World Health Organization's report on suicide: a fundamental step in worldwide suicide prevention. Crisis, 35(5), 289–291. https://doi.org/10.1027/0227-5910/a000293

Healthline (n.d.). Origin. [Website]. Retrieved from: https://www.healthline.com/health/depression/facts-statistics-infographic#1

Tabular data: 1K rows x 6 columns. suicide rate of each gender in countries around the world in 2000, 2010, 2015 and 2016.(https://apps.who.int/gho/data/view.main.MHSUICIDEv)

Tabular data: 1K rows x 63 columns. GDP of global countries from 1960-2018 (https://data.worldbank.org/indicator/NY.GDP.MKTP.CD)